

White Paper Report

Report ID: 103902

Application Number: HW5002111

Project Director: Roger Bagnall (roger.bagnall@nyu.edu)

Institution: New York University

Reporting Period: 9/1/2011-8/31/2012

Report Due: 11/30/2012

Date Submitted: 11/26/2012

Planning a Digital Corpus of Greek and Latin Literary Papyri

Rodney Ast (University of Heidelberg) and Roger S. Bagnall (New York University)

Project Aims

Since the end of the 19th century, Egypt and, to a lesser extent, other places in the Mediterranean have yielded thousands of papyri, most written in Greek and others in Egyptian, Latin, Arabic and other languages. The majority of them contain documentary texts concerned with public and private affairs in Egypt under Ptolemaic, Roman, and Islamic rule, but an estimated 15% are ancient books of varied content, such as epic, ancient drama, lyric poetry, historical and oratorical works, medical treatises, lexicographical texts, etc. Some works are known because copies have been preserved or cited in later manuscript traditions, but others survive only in papyri, for example, the poetry of Sappho or the Aristotelian *Constitution of the Athenians*, while still others were written by authors whose identity is unknown to us. Literary texts, particularly those by identifiable authors, have tended to elicit more enthusiasm among classicists than the average document, yet documentary papyri have been better represented in the digital sphere, thanks in large part to the existence of an online database of Greek and Latin documentary texts known as the Duke Databank of Documentary Papyri (DDbDP). The DDbDP comprises searchable transcriptions of nearly the entire corpus of extant Greek and Latin documentary papyri. While some literary papyri can be found in the Thesaurus Linguae Graecae (TLG), and others are available via more targeted projects, such as the Catalogue of Paraliterary Papyri (CPP) and Thesaurus Herculanensium Voluminum (THV), no larger corpus of texts exists in digital form. For that reason, in the Spring and Summer 2012 we conducted two workshops centered around a simple question: What would it take to create the infrastructure for a digital corpus of literary papyri similar to the DDbDP but reflective of the different source material and responsive to the particular needs of scholars who deal with it?

The time for developing a digital corpus of literary papyri is ripe. Over the course of a three-year, Mellon-funded project entitled Integrating Digital Papyrology (IDP), the DDbDP was converted to an XML format known as EpiDoc and given a new home at papyri.info. The databank was also equipped with an editorial system called the Papyrological Editor (PE; formerly known as SoSOL) that allows papyrologists to contribute new documents and propose emendations to existing texts online (since its release in 2010, over 1000 new texts have already been added). The purpose of these developments was to make the Databank easier to manage by opening it up to the larger scholarly community, and open-source tools were employed in order to make the technology available for others to create analogous, interoperable digital resources. Our aim has been to capitalize on these advancements and create a searchable and editable corpus of literary texts that will allow scholars to identify, classify, and study genres and authors more easily. We intend for this corpus to be searchable within the PN alongside documents in the DDbDP, in order, in part, to narrow the traditional divide between documentary and literary studies especially in areas like linguistics, social history, medicine, religious studies, etc. What follows is a brief report of the results of our two meetings, to which we invited specialists in literary papyrology and digital programs (see the list of participants below) who helped us define the scope and technical requirements for this project.

Results of Planning Workshops

The aim of the two highly successful planning meetings was twofold: a) to take a broad view of existing and planned digital projects not only in papyrology but also in other areas of ancient studies in order to identify ways in which existing technologies could be leveraged for literary papyri and to prioritize development of new technologies in light of these other planned initiatives; b) to sketch an "ideal" corpus of literary papyri by tapping into the wishes and current pilot projects of specialists in the field.

With IDP winding down there is at least one major initiative in the works, currently at the planning stage. The American Philological Association (APA) has received a grant from Mellon (through mid-2013) to sketch plans for a full-scale database of Latin literature (from the earliest examples through the medieval period) built on IDP technology.

In addition to the APA project, Mellon recently convened a meeting about the future of digital projects related to ancient Greek and has subsequently funded a planning project (Integrating Digital Greek) led by Joshua Sosin. Mellon's commitment is to a future in which there is a more open digital corpus of Greek literature; this is not so much about fees for service as it is about data availability for other interests. The place of TLG within this future remains uncertain. This has bearing on a literary papyrology project since TLG has obvious points of relevance (searchable papyrological texts and later witnesses to textual traditions found in extant literary papyri). One aspect of a literary project would be to explore with TLG mutually beneficial modes of interaction.

DCLP Basic Components

The two workshops generated the following clear statement about the DCLP's mission. It

- Will exist within space of the PE/PN community and will be based on IDP technology
- Will use TM-LDAB as basic metadata source for published texts, to be supplemented by genre-specific projects (e.g., Herculaneum and Parma Medical Project)
- Will be created by multiple teams for various components, using common standards and technology, with a collaborative structure to be determined
- Will develop in close relationship with Latin databank and Greek tooling initiatives launched with Mellon support and seek to minimize separate software development
- At a basic level can be created with existing PE/PN tools, but requires some up-front development work as well as broader tool creation of Mellon initiatives
- At a basic level can be created with existing PE/PN tools, but requires some up-front development work as well as broader tool creation of Mellon initiatives
- Begins with Parma, Herculaneum, Ductus, and CPP projects but needs to enable other spontaneous communities as interest develops in near term
- Will enable local identity ("branding") of projects via extraction functions rather than maintenance of separate repositories

DCLP Specific Enhancements

The participants enumerated a large list of desiderata for the DCLP:

- Creation of publication metadata module for PE; crosswalk of LDAB to module and report on lessons learned; extension of metadata mask to accommodate a range of enhancements
- Some EpiDoc-TEI and Leiden + development
- Begins with Parma, Herculaneum, and Ductus projects but needs to enable other spontaneous communities as interest develops in near term
- Will enable local identity ("branding") of projects via extraction functions rather than maintenance of separate repositories
- Begins with Parma, Herculaneum, and Ductus projects but needs to enable other spontaneous communities as interest develops in near term
- Crosswalk of CPP texts into PE to diagnose issues

Long-Term Development (Non-DCLP)

- Ability to share work in progress widely
- Uploading of images by individual contributor
- Discovery of work in progress/ cries for help
- Discussions with TLG about use of texts/stable URLs
- Openness of data to use by external tools/platforms

Discussions with papyrologists identified a range of coding enhancements, some of which we would intend as DCLP deliverables; we view the others as general improvements better left to the other pending projects

Coding

I. DCLP

- Expanded symbols, abbreviations, punctuation (e.g., dicola,) signs (coronis, paragraphos, etc.)
- Markup of known hands
- Pin LDAB/Meta- data to text divs. such as fragments or columns
- Allow for anomalies (e.g. certain features of Herculaneum papyri) via corpus-specific LDAB fields
- Resp attribute for ancient hands (and modern editors?)
- Diplomatic and normal texts
- Meter
- Dialect information (XML lang needed for this)
- Eisthesis + ecthesis: physical form vs meaning
- Positioning + contents of fragments (sottoposti etc.)
- Apparatus support for
 - Distinguishing witnesses (as in P.Herc. or multiple copies) & "Lematic" representation of witnesses/readings
 - Orthographic errors/variants
 - Broader textual tradition/management of mult. editions

-Consistent punctuation of subunits

- Line numbers of literary works (or other standard citation)
- Relationship of front+back sides

II. Long-term

- Accommodation of tabular data, diagrams, pictures, drawings
- Musical notation
- Scholia in various formats, marginalia
- Colometry
- Isopsephism
- Dependency relationships in lemmas
- Better representation (semantically aware) of blank spaces (blank space to indicate divisions--e.g., caesura)
- Ancient commentaries: accommodation of lemmata, sublemmata, quotations of ancient authors, abbreviations
- Biblio links live in every part of record

Needs: display & user

I. DCLP

- More use of drop-down menus and autofill
- Additional browse facets:

-author and work

-meter

-line number or range

-dialect

-genre/type

-lemmatization

- Standardization of edition info
- Bold face for the standard/cited/base edition in display

II. Long-term

- Searchability of non-verbal content like signs
- Searchability of text within text (e.g., quotations)
- Sorting & counting of results; visual display
- Linkage of text to line in image
- Export of results to CSV
- Bookmark results option; search save function
- Multi-lingual lexicon to support cross-language searching of metadata & biblio.
- Pattern matching across texts (quoting, intertextuality, etc.)
- Network analysis: or external product?
- Use of number server tables (?) to escape citation format issues like forms of papyrus abbreviations

The University of Heidelberg and New York University take this opportunity to express their thanks to the DFG and NEH for supporting this planning work. As a result, we have obtained

a clear idea of both what the field of literary papyrology needs in terms of both coding of textual data and metadata and functionalities required in a user interface. We are using the results of this study to prepare an implementation proposal that adopts these recommendations and seeks to carry out a first phase of work with full cooperation with the other projects, both papyrological and literary, currently underway.

In conclusion, we would say that we see the work of this project as reflecting broader trends in the study not only of papyri but of all kinds of ancient and medieval textual supports: the physical and visual characteristics of objects with writing, whether papyri, potsherds, or stones, have come to occupy a much more prominent role in our studies. In part this development reflects the vitality of the history of the book as a field of study in recent decades; in part the greater integration of archaeological evidence into documentary fields; and in part the increasingly widespread availability of digital images since the mid-1990s. We believe that the general direction of the scholarly needs identified in this study is applicable to areas in ancient and medieval studies other than literary papyrology, and the provision of those needs should and, we believe, will be carried out in collaboration with these broader communities.

APPENDIXES

1. List of Participants

Meeting 1; April 2 - 3, 2012
NYU's Institute for the Study of the Ancient World (ISAW)
New York

- Bridget Almas (Tufts)
- Rodney Ast (Heidelberg)
- Roger Bagnall (ISAW)
- Ryan Baumann (Univ. of Kentucky)
- Hugh Cayless (NYU)
- Raffaella Cribiore (NYU)
- Mark Depauw (Leuven)
- Tom Elliott (NYU)
- Alexander Jones (NYU)
- Julia Lougovaya (Heidelberg)
- AnneMarie Luijendijk (Princeton)
- Joshua Sosin (Duke)

Meeting 2; June 18 - 19, 2012
University of Heidelberg

- Isabella Andorlini (Parma)
- Rodney Ast (Heidelberg)
- Roger Bagnall (NYU)

- Ryan Baumann (Univ. of Kentucky)
- Hugh Cayless (NYU)
- James Cowey (Heidelberg)
- Mark Depauw (Leuven)
- Holger Essler (Würzburg)
- Bill Furley (Heidelberg)
- Jürgen Hammerstaedt (Cologne)
- Andrea Jördens (Heidelberg)
- Carmen Lanz (Heidelberg)
- Julia Lougovaya (Heidelberg)
- John Lundon (Cologne)
- Alberto Nodar (Barcelona)
- Andreas Schwab (Heidelberg)
- Joshua Sosin (Duke)

2. Papyrological Acronyms used above

APIS	Advanced Papyrological Information System
CPP	Corpus of Paraliterary Papyri
HGV	Heidelberg Gesamtverzeichnis
IDP	Integrating Digital Papyrology project
ISAW	Institute for the Study of the Ancient World, NYU
LDAB	Leuven Database of Ancient Books
PE/PN	Papyrological Editor/Papyrological Navigator
TLG	Thesaurus Linguae Graecae
TM	Trismegistos